# Raising the standard of published systematic reviews

A case study from chemical risk research

Paul Whaley

23 May 2018

# About me

- Researcher at Lancaster University and the Evidence-Based Toxicology Collaboration at Johns Hopkins BSPH
- Background in environmental health advocacy and science communication
- Introduced to systematic reviews as gold-standard approach to evidence synthesis in early 2010
- Associate Editor for Systematic Reviews at *Environment International* (IF 7.088) – first specialist EH SR editor
- The “frameworks guy”: systematic approaches to evidence surveillance and synthesis; critical appraisal tools; codes of practice; research quality management

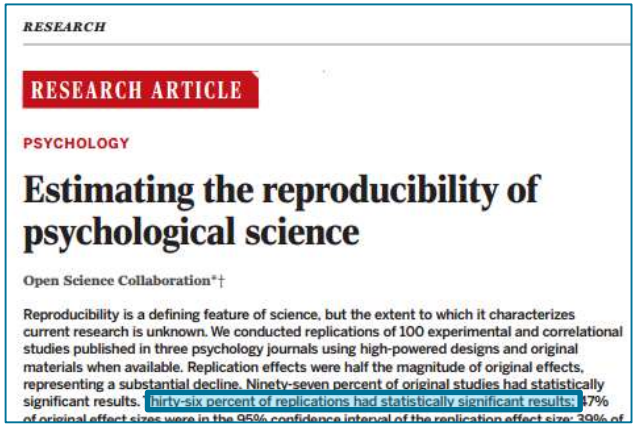
# Today's presentation

- Reproducibility issues in chemical risk assessment as a driver of interest in systematic review methods
- Uptake of SR methods
- Challenges we are seeing (poor quality SRs)
- How we are addressing these challenges at *Environment International*
- Implications for you as potential submitting authors and conductors of systematic reviews

# A “reproducibility crisis” in primary research

VIEWPOINT

## The Proposal to Lower *P* Value Thresholds to .005

*RESEARCH*

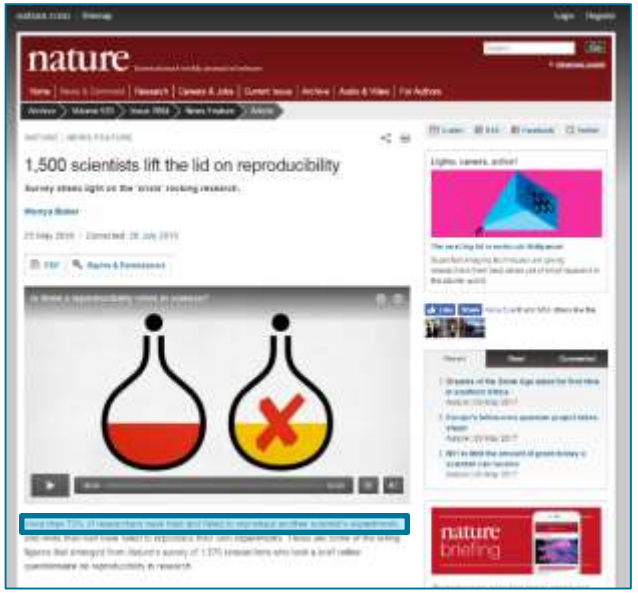
RESEARCH ARTICLE

PSYCHOLOGY

## Estimating the reproducibility of psychological science

Open Science Collaboration*†

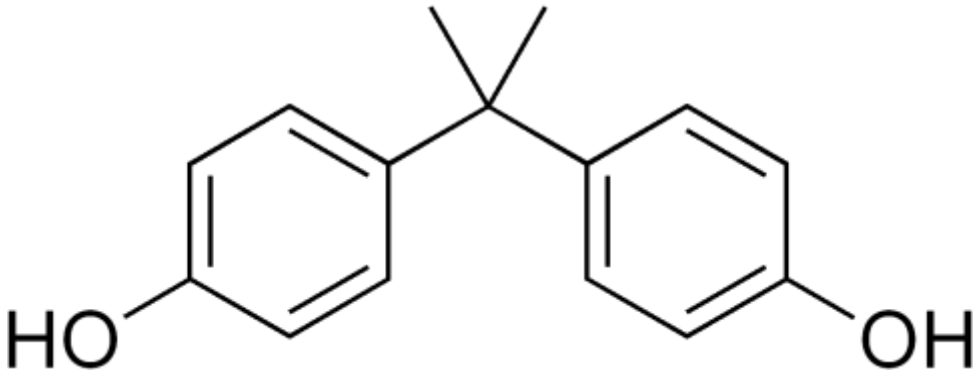
Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47%



# Chemical risk assessment

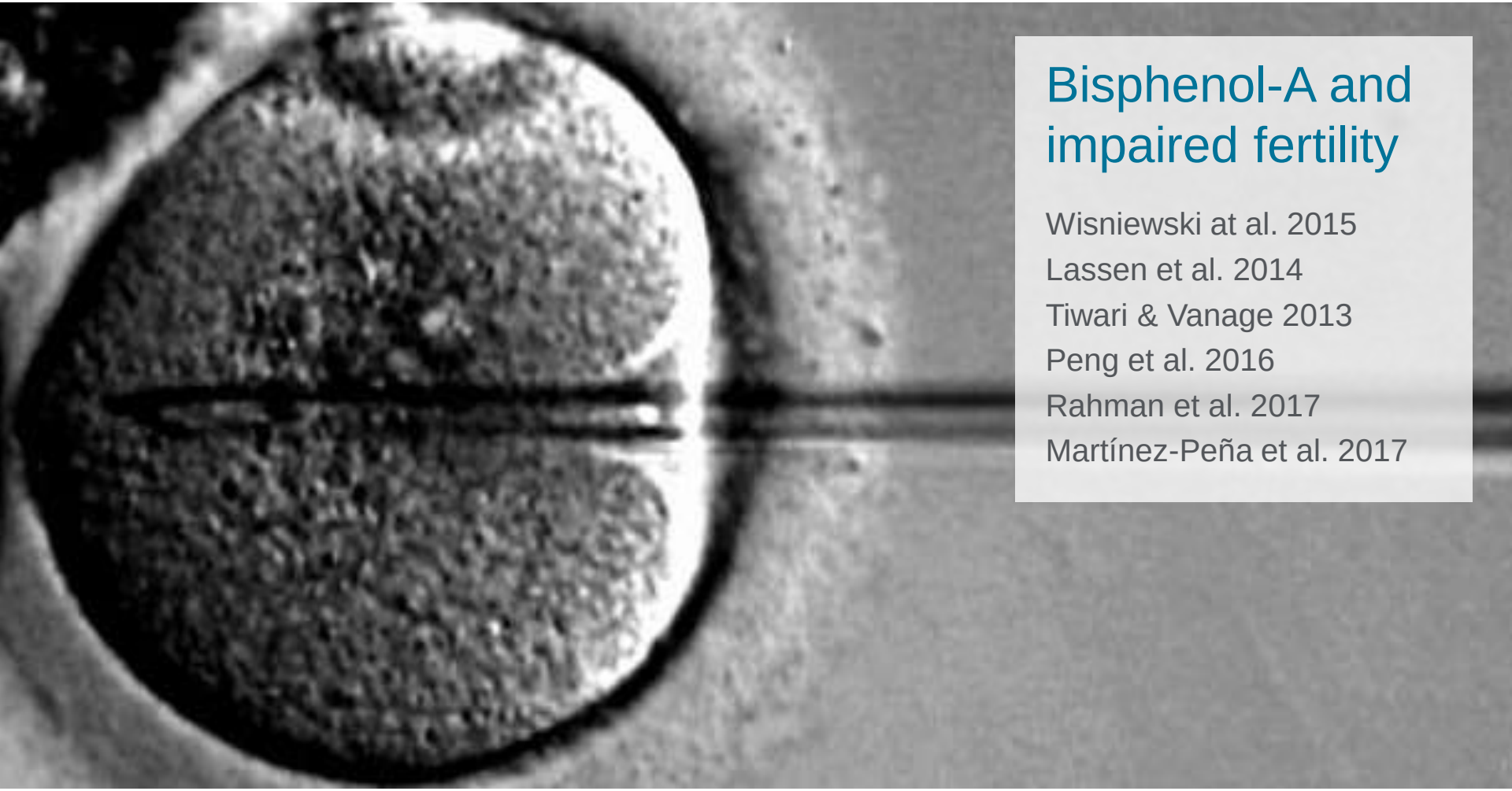
- Making sense of complex and contradictory evidence about health risks posed by exposure to chemical substances

# Reproducibility crisis in chemical risk assessment



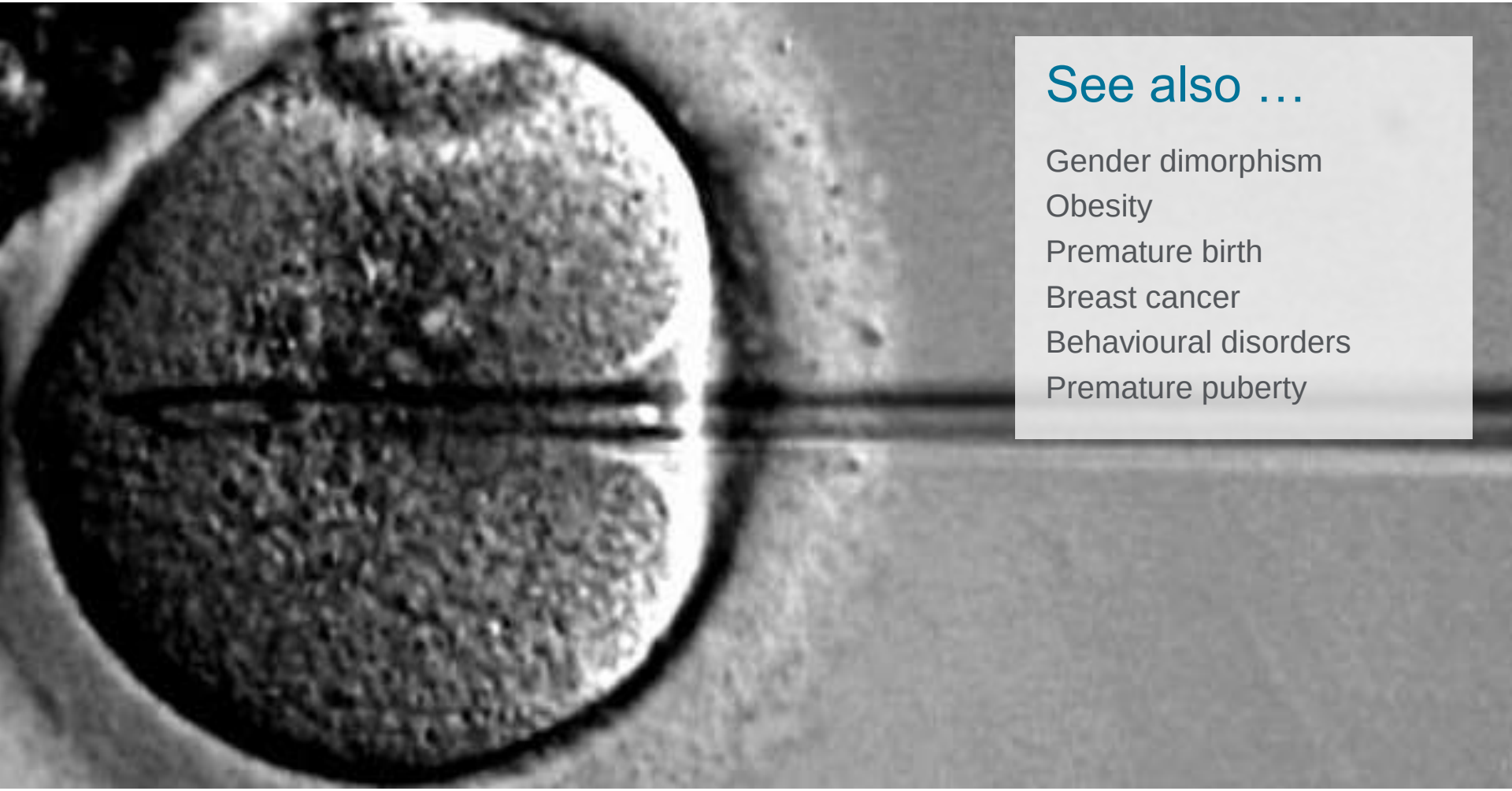

Bisphenol-A


POWER



# Bisphenol-A and impaired fertility

Wisniewski at al. 2015
Lassen et al. 2014
Tiwari & Vanage 2013
Peng et al. 2016
Rahman et al. 2017
Martínez-Peña et al. 2017



## See also …

Gender dimorphism
Obesity
Premature birth
Breast cancer
Behavioural disorders
Premature puberty

efsa
European Food Safety Authority

Karolinska Institutet

Public Health England

International Agency Research on Cancer

World Health Organization

FDA

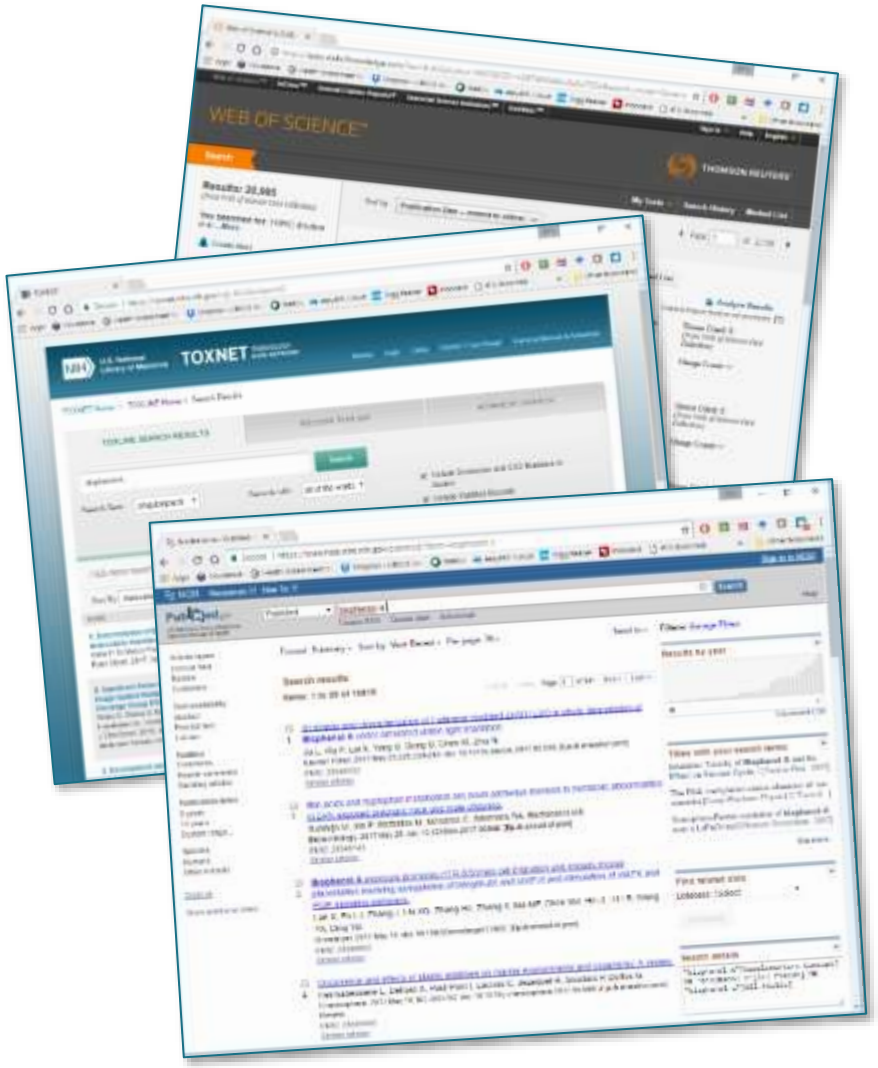
…effects have been demonstrated for BPA [at] levels **10–10,000x lower** than the current LOAEL of 50 mg/kg/day
Vandenberg et al. 2014


…**no health concern** for any age group from dietary exposure
EFSA 2015

…a TDI for BPA has to be **0.7 µg/kg bw/day** or lower to be sufficiently protective
National Food Institute, Denmark 2015

…a **potential risk to the unborn children** of exposed pregnant women [relating to] a change in the structure of the mammary gland
ANSES 2013

# Same evidence, different conclusions




?

…**no health concern** for any age group from dietary exposure

…a TDI for BPA has to be **0.7 µg/kg bw/day** or lower to be sufficiently protective
National Food Institute

…effects have been demonstrated for BPA [at] levels **10–10,000x lower** than the current LOAEL of 50 mg/kg/day
Vandenberg et al. 2014

# Solving the problem with systematic review methods

- Accelerating uptake since I started working on this in 2010

NTP
National Toxicology Program
U.S. Department of Health and Human Services

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

GRADE

World Health Organization

efsa
European Food Safety Authority

UCSF
University of California San Francisco

Program on Reproductive Health and the Environment

ebtc
Evidence-based Toxicology Collaboration

# Rapid growth in publication of SRs


TITLE: ("systematic review"); Refined by: WEB OF SCIENCE CATEGORIES: ( TOXICOLOGY ) AND [excluding] WEB OF SCIENCE CATEGORIES: ( PHARMACOLOGY PHARMACY ); Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, IC.

# But we have a problem with quality

- 8989 PubMed records tagged by 2004 as “systematic review” yet actual number of stringently-defined SRs was ~2500 (Moher et al. 2007)
- Most published SRs have major flaws in conduct and reporting (Page et al. 2016)
- ~3% of manuscripts are “decent and clinically useful” (Ioannidis 2016)
- Our own pilot data shows serious omissions in reporting of 19 of 25 SRs published in the top environmental health journals through 2014-2015, before we even look at the validity of the actual methods used
- Fundamental errors mean a lot of effort is being put into projects which are not fit for purpose

# My job as an editor

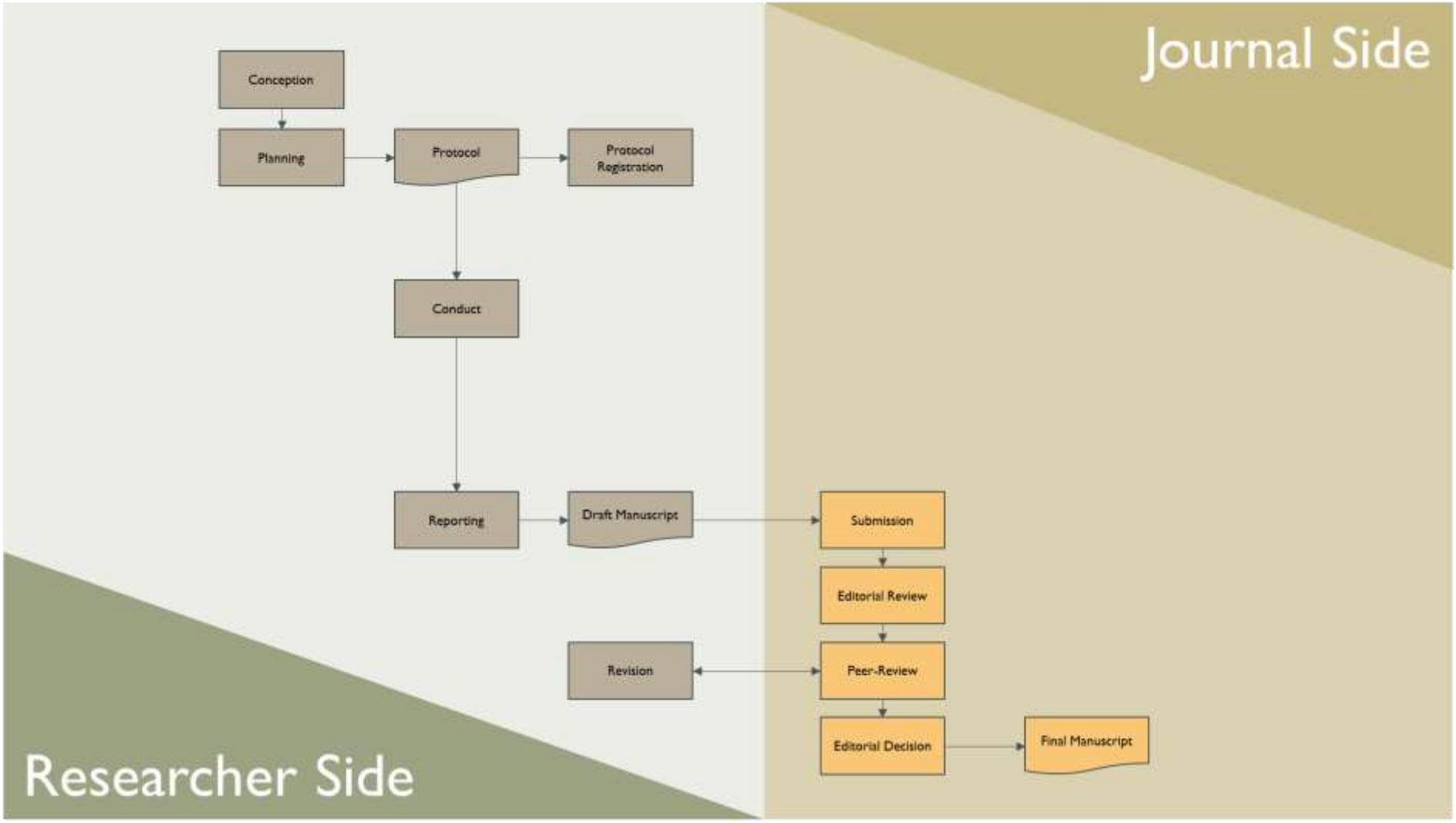
- What can I do at our journal to ensure each SR we publish is fit for purpose?
  - Asks an important question
  - Is truthful
  - Includes all information about methods and results, such that a reader can appraise the validity of the SR's findings and assess its relevance to their decision-making context
- Gatekeeper and midwife strategies for ensuring we publish high-quality research
- Implications for you as researchers

# EDITOR AS GATEKEEPER

Enforcement of reporting standards

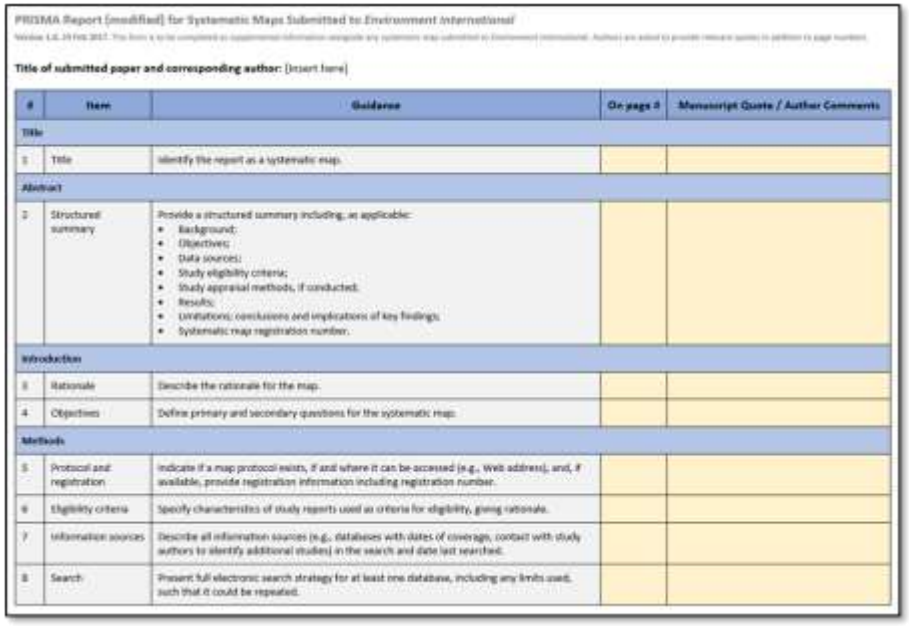
Editorial triage

Making best use of peer-review


Journal Side
Conception
Planning
Protocol
Protocol Registration
Conduct
Reporting
Draft Manuscript
Submission
Editorial Review
Revision
Peer-Review
Editorial Decision
Final Manuscript
Researcher Side

# Enforcement of reporting standards

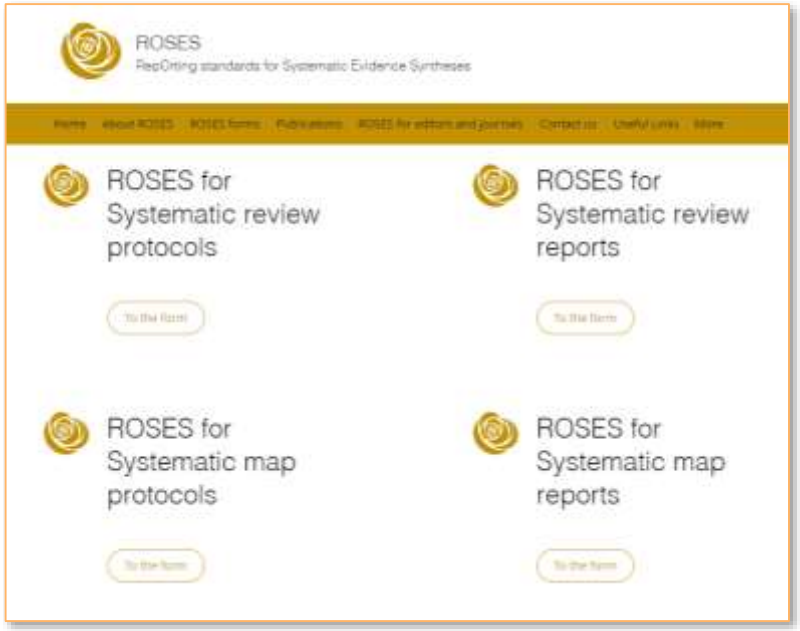
- Option of PRISMA (Moher et al. 2009) or ROSES (Haddaway et al. 2018)
- Submission of PRISMA or ROSES report as supplemental information is compulsory
- Useful quick check on basic standards



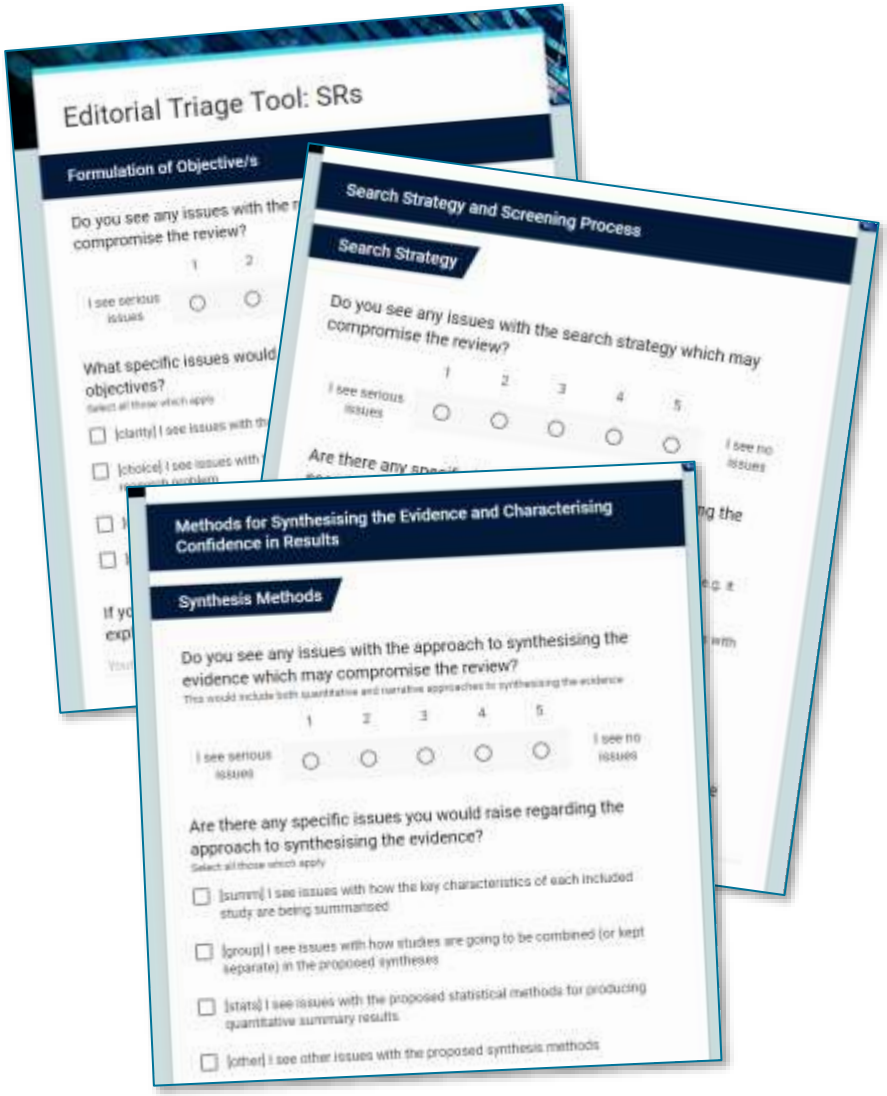
PRISMA Report (modified) for Systematic Maps Submitted to *Environment International*

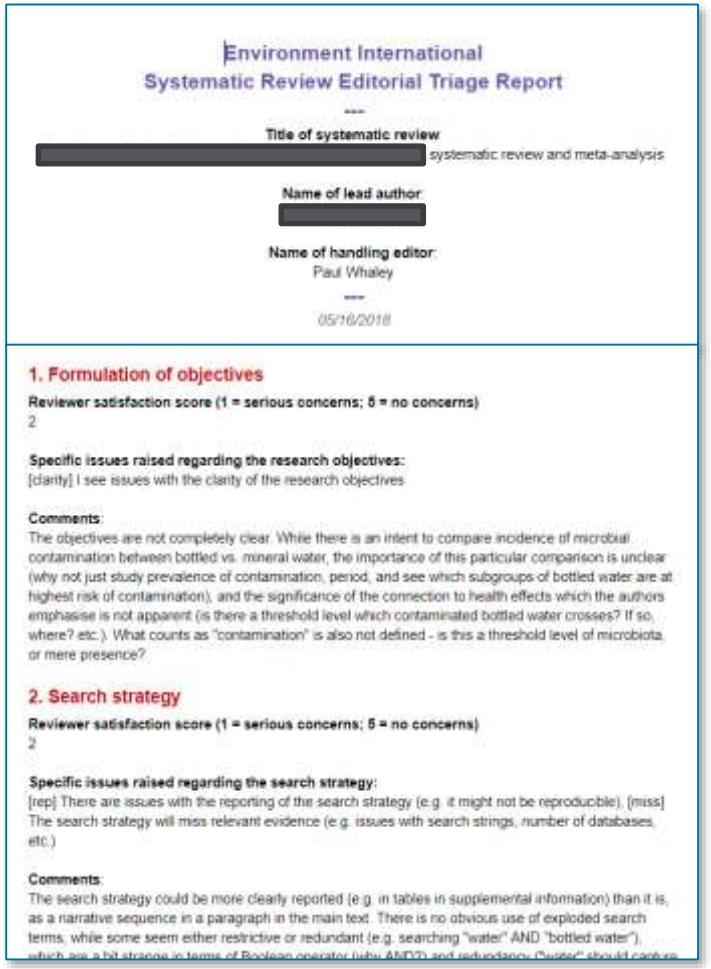
**Title of submitted paper and corresponding author:** (Insert here)

| # | Item | Guidance | On page # | Manuscript Quote / Author Comments |
|---|---|---|---|---|
| **Title** | | | | |
| 1 | Title | Identify the report as a systematic map. | | |
| **Abstract** | | | | |
| 2 | Structured summary | Provide a structured summary including, as applicable:<br>• Background;<br>• Objectives;<br>• Data sources;<br>• Study eligibility criteria;<br>• Study appraisal methods, if conducted;<br>• Results;<br>• Limitations; conclusions and implications of key findings;<br>• Systematic map registration number. | | |
| **Introduction** | | | | |
| 3 | Rationale | Describe the rationale for the map. | | |
| 4 | Objectives | Define primary and secondary questions for the systematic map. | | |
| **Methods** | | | | |
| 5 | Protocol and registration | Indicate if a map protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | | |
| 6 | Eligibility criteria | Specify characteristics of study reports used as criteria for eligibility, giving rationale. | | |
| 7 | Information sources | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | | |
| 8 | Search | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | | |



# Editorial triage reports



Environment International
Systematic Review Editorial Triage Report

---

**Title of systematic review**
[redacted] systematic review and meta-analysis

**Name of lead author**:
[redacted]

**Name of handling editor**:
Paul Whaley

---

*05/10/2018*

## 1. Formulation of objectives

**Reviewer satisfaction score (1 = serious concerns; 5 = no concerns)**
2

**Specific issues raised regarding the research objectives:**
[clarity] I see issues with the clarity of the research objectives

**Comments**:
The objectives are not completely clear. While there is an intent to compare incidence of microbial contamination between bottled vs. mineral water, the importance of this particular comparison is unclear (why not just study prevalence of contamination, period, and see which subgroups of bottled water are at highest risk of contamination), and the significance of the connection to health effects which the authors emphasise is not apparent (is there a threshold level which contaminated bottled water crosses? If so, where? etc.). What counts as "contamination" is also not defined - is this a threshold level of microbiota or mere presence?

## 2. Search strategy

**Reviewer satisfaction score (1 = serious concerns; 5 = no concerns)**
2

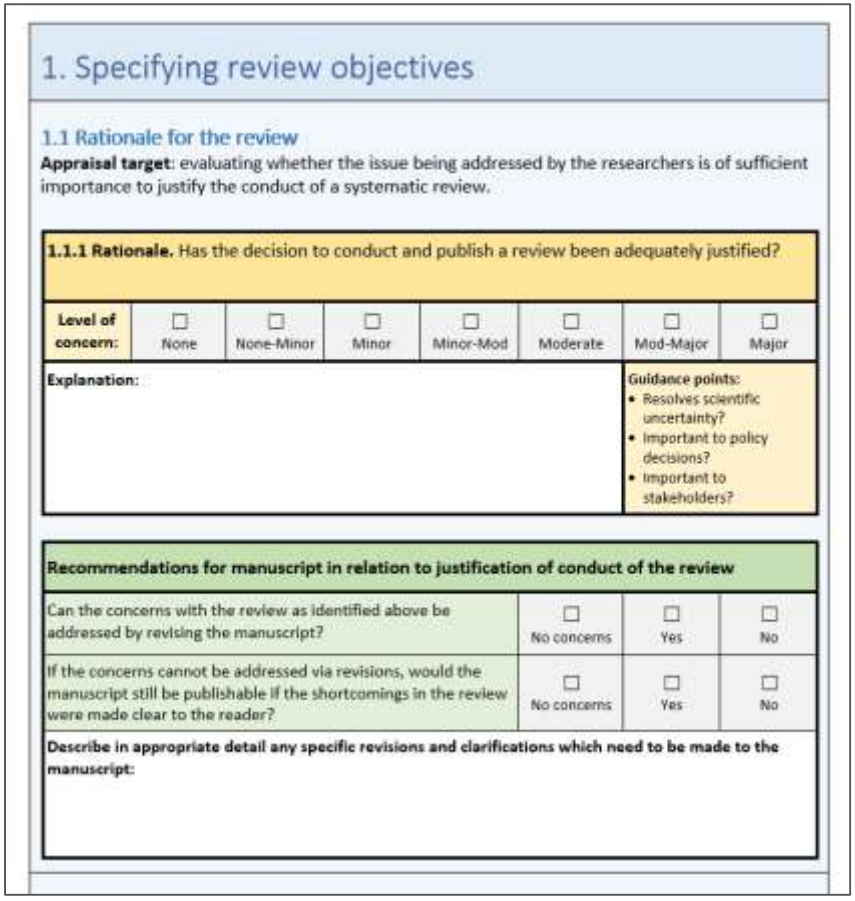
**Specific issues raised regarding the search strategy:**
[rep] There are issues with the reporting of the search strategy (e.g. it might not be reproducible), [miss] The search strategy will miss relevant evidence (e.g. issues with search strings, number of databases, etc.)

**Comments**:
The search strategy could be more clearly reported (e.g. in tables in supplemental information) than it is, as a narrative sequence in a paragraph in the main text. There is no obvious use of exploded search terms, while some seem either restrictive or redundant (e.g. searching "water" AND "bottled water"), which are a bit strange in terms of Boolean operator (why AND?) and redundancy ("water" should capture

# Improved peer-review

- Target of 4 reviewers per submission
  - 2 topic experts
  - 2 methods experts
- Peer-review facilitation tool
  - Testing a Google Forms tool similar to Triage tool
  - Building CREST-SR for full-blooded implementation

Whaley et al. “A Tool for Critical Appraisal of Evidence Syntheses in Toxicology: Systematic Reviews (CREST-SR)” Under development



## 1. Specifying review objectives

### 1.1 Rationale for the review

**Appraisal target**: evaluating whether the issue being addressed by the researchers is of sufficient importance to justify the conduct of a systematic review.

**1.1.1 Rationale.** Has the decision to conduct and publish a review been adequately justified?

| Level of concern: | ☐ None | ☐ None-Minor | ☐ Minor | ☐ Minor-Mod | ☐ Moderate | ☐ Mod-Major | ☐ Major |
|---|---|---|---|---|---|---|---|

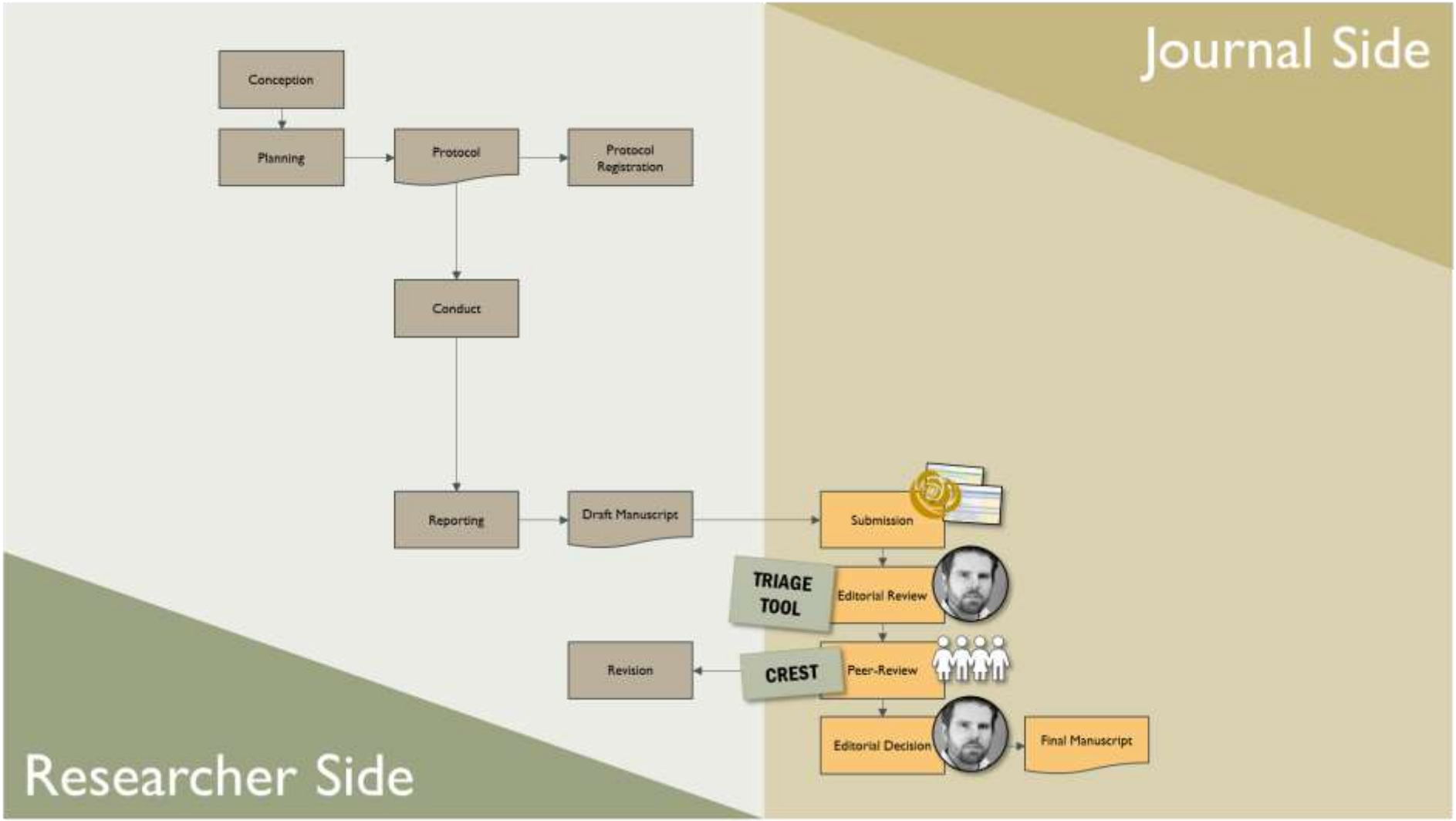
**Explanation:**

**Guidance points:**
- Resolves scientific uncertainty?
- Important to policy decisions?
- Important to stakeholders?

**Recommendations for manuscript in relation to justification of conduct of the review**

| | | | |
|---|---|---|---|
| Can the concerns with the review as identified above be addressed by revising the manuscript? | ☐ No concerns | ☐ Yes | ☐ No |
| If the concerns cannot be addressed via revisions, would the manuscript still be publishable if the shortcomings in the review were made clear to the reader? | ☐ No concerns | ☐ Yes | ☐ No |

**Describe in appropriate detail any specific revisions and clarifications which need to be made to the manuscript:**


Journal Side
Researcher Side
Conception
Planning
Protocol
Protocol Registration
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript

# Progress so far?

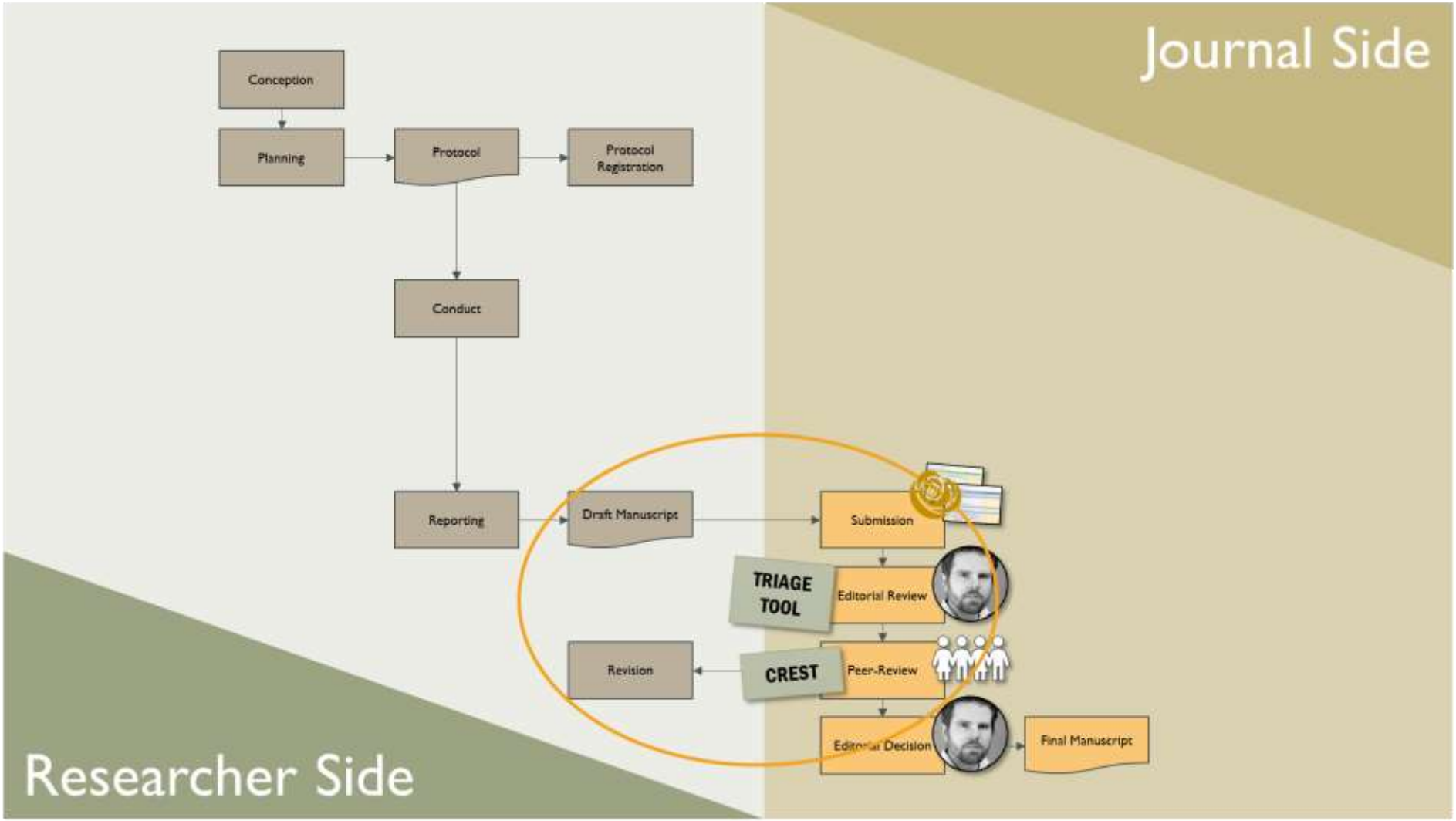
- 46 of 67 submissions rejected since using EVISE (~18 months)
  - 10 in process, 10 sent to production, one declined resubmission
  - 6 SRs, one SM, 2 commentaries, one correspondence
  - Only 3 SRs rejected post peer-review, 43 pre peer-review
- Hopefully that means we are at least filtering out the SRs which are not fit for purpose

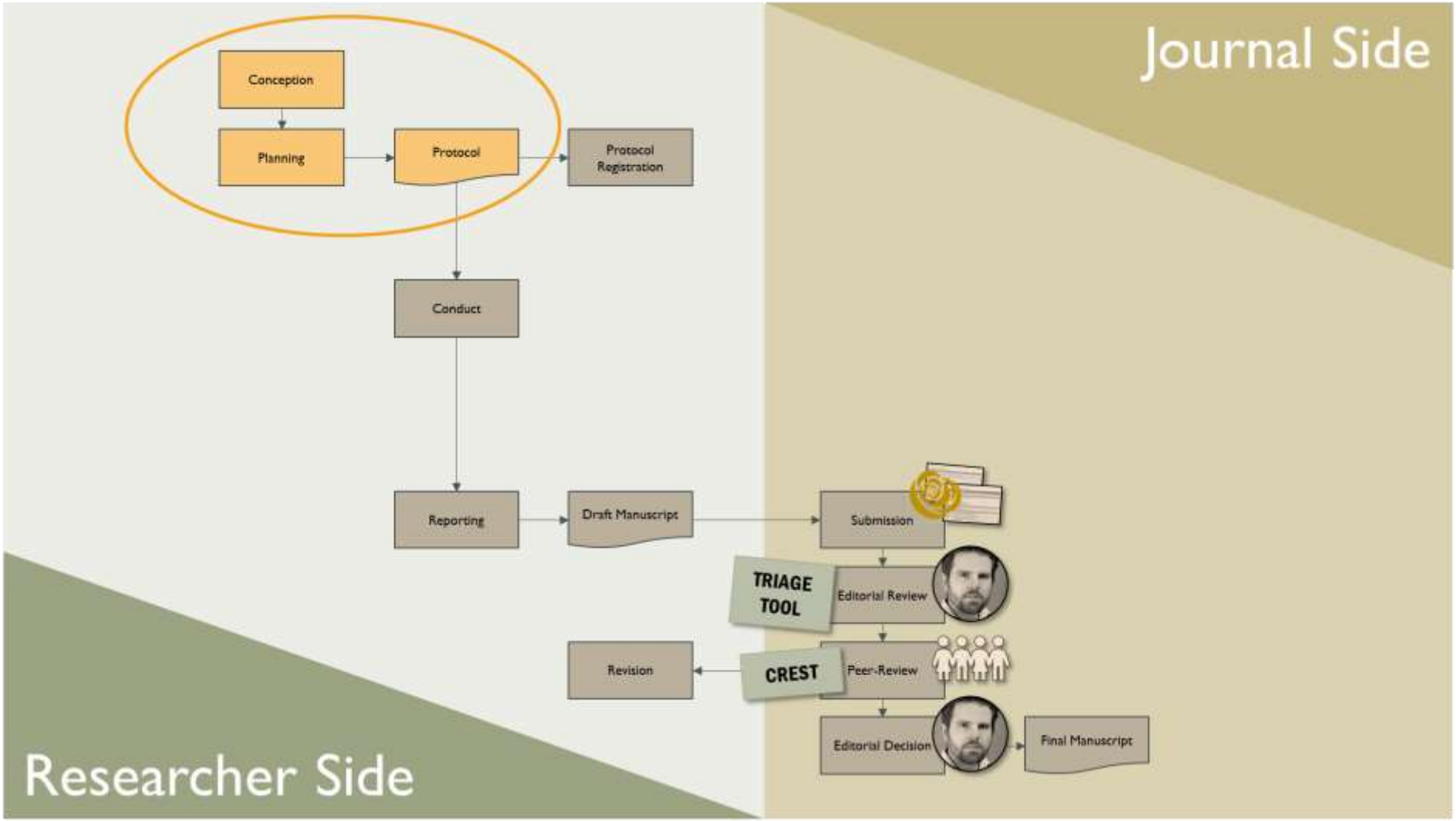
# Is it really progress?

- We are mainly getting low-quality systematic reviews long after it's too late for the authors to address major issues (43 of 46 rejections are at desk; 2 years of work rejected in 2 minutes)
  - Objectives lacking research value and/or focus
  - Insensitive search strategies
  - Inappropriate inclusion criteria
  - Inadequate or non-existent risk of bias assessment methods
  - Unstructured, unsystematic interpretation of strength of evidence
- We are making sure readers aren't receiving misleading research (at least through our own journal) but could do much more to help submitting authors develop high-quality manuscripts

# EDITOR AS MIDWIFE

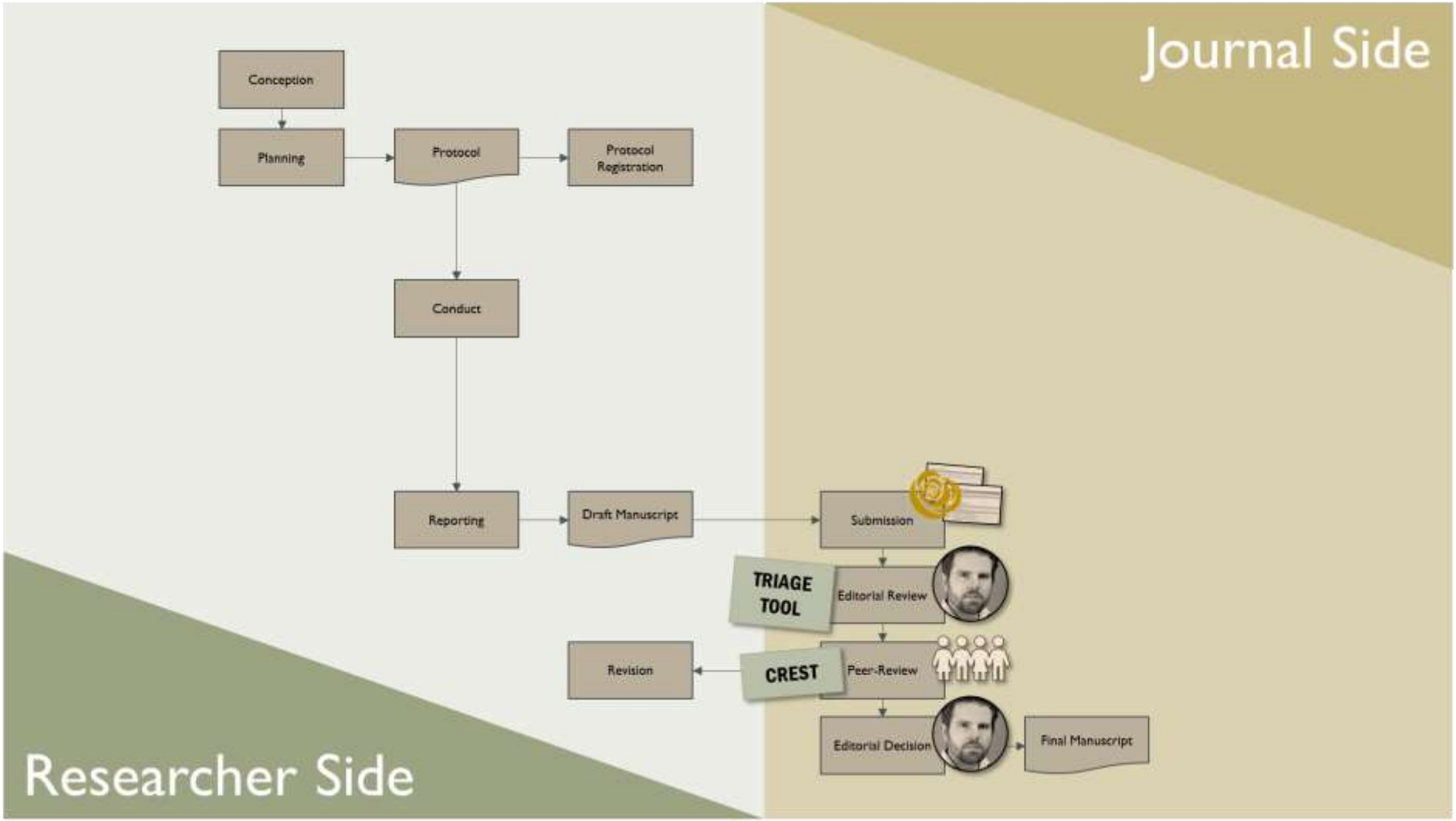
Rethinking the SR workflow and submission process

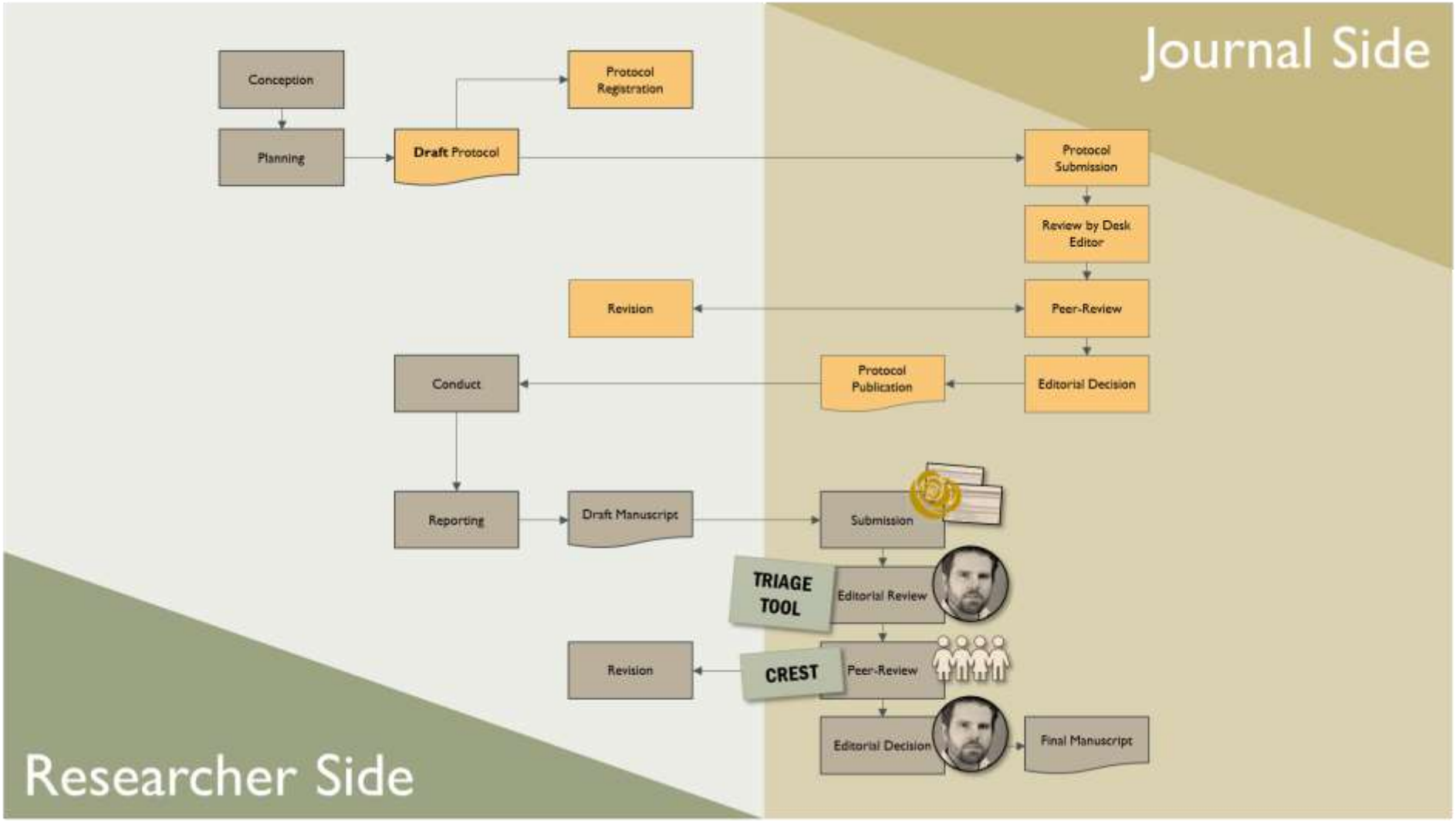

Journal Side
Researcher Side
Conception
Planning
Protocol
Protocol Registration
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript

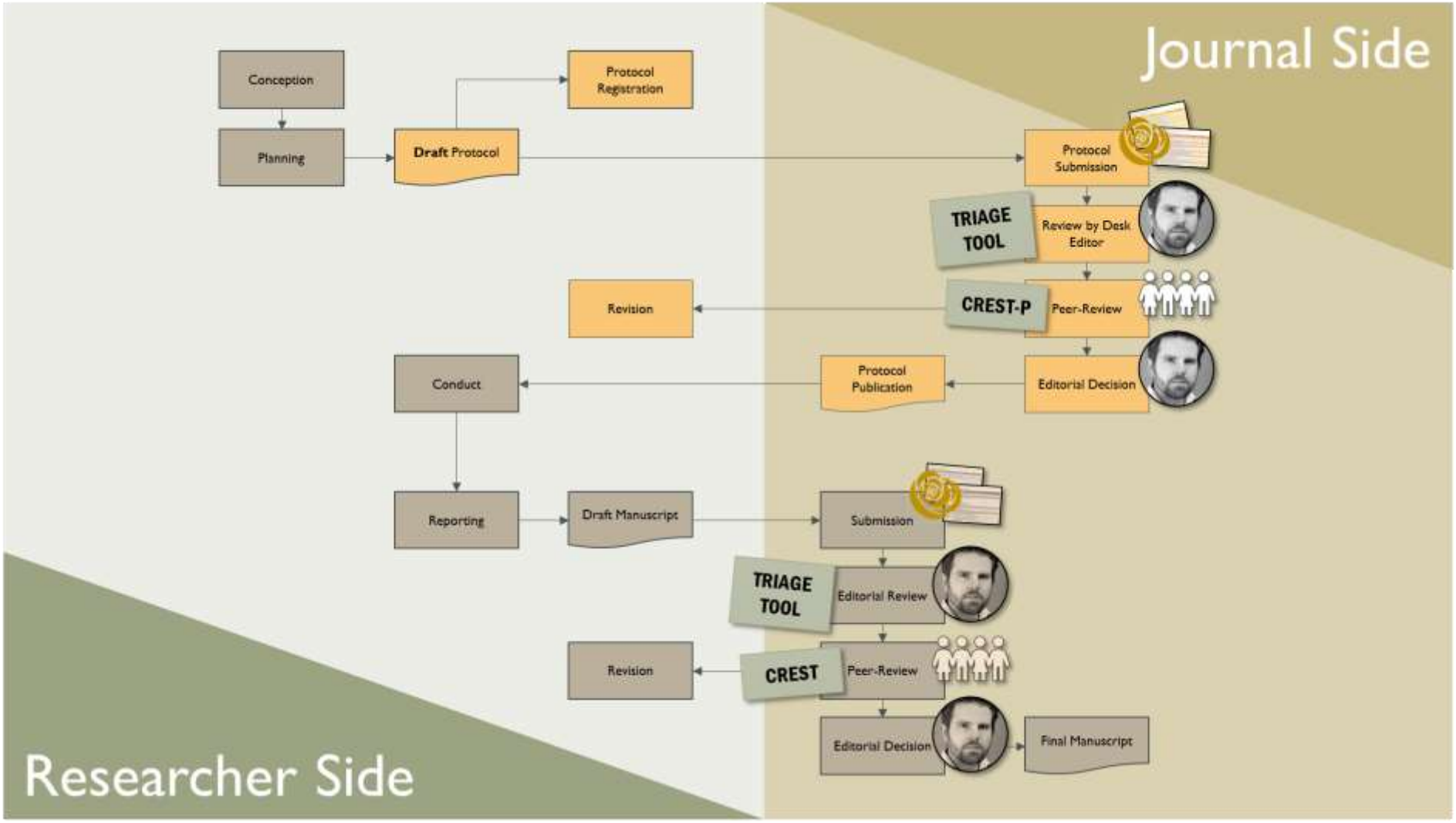

Journal Side
Researcher Side
Conception
Planning
Protocol
Protocol Registration
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript

# The solution: accept protocol submissions

- *Environment International* counts protocols as full publications
- First environmental health journal to do this
- Opens up multiple opportunities for editorial interventions


Journal Side
Conception
Planning
Protocol
Protocol Registration
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
Revision
CREST
Peer-Review
Editorial Decision
Final Manuscript
Researcher Side


Journal Side
Researcher Side
Conception
Planning
Draft Protocol
Protocol Registration
Protocol Submission
Review by Desk Editor
Peer-Review
Editorial Decision
Protocol Publication
Revision
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript


Journal Side
Researcher Side
Conception
Planning
Draft Protocol
Protocol Registration
Protocol Submission
TRIAGE TOOL
Review by Desk Editor
CREST-P
Peer-Review
Editorial Decision
Protocol Publication
Revision
Conduct
Reporting
Draft Manuscript
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript

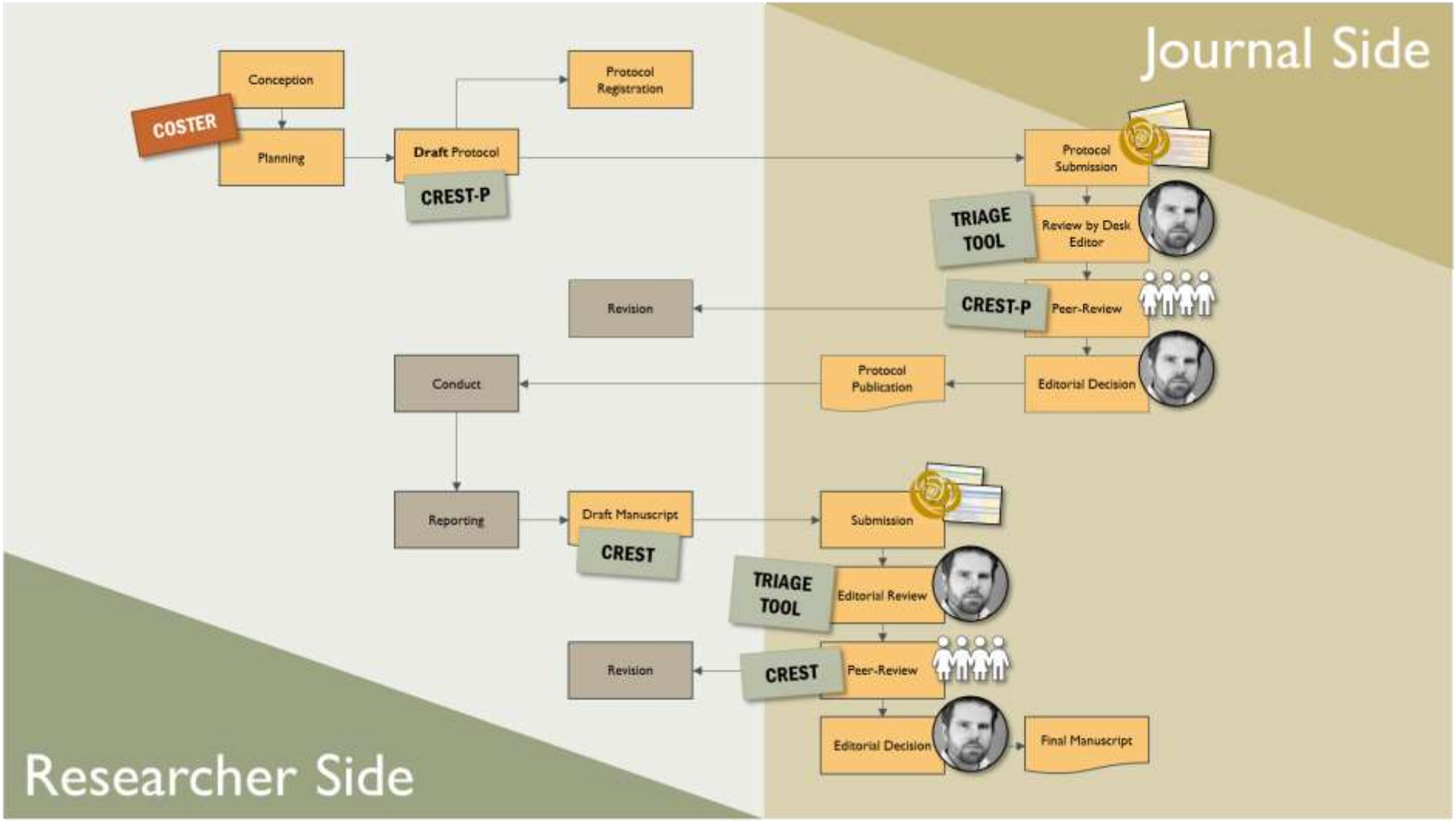
# Final piece of the puzzle

- "Recipe-book" for what researchers ought to do, to maximise chance of producing a fit-for-purpose systematic review
- Developing a tool called COSTER – 70 provisions across 8 stages of conducting a systematic review
- Makes explicit the required processes for fulfilling the criteria of e.g. PRISMA or ROSES, and for critical appraisal tools such as CREST

Step 3: Screening Evidence for Inclusion

| Proposed Wording | Comments | Notes for explanation / elucidation document |
|---|---|---|
| | | |
| 3.1 Screening of each piece of evidence for inclusion to be conducted by at least two people working independently, with an appropriate process (e.g. third party arbitration) for identifying and settling disputes. | | |
| 3.2 Document decisions in enough detail to allow presentation of the results of the screening process in a PRISMA flow chart. | | |

Whaley et al. "A Code of Practice for Conduct of Systematic Reviews in Toxicology and Environmental Health Research (COSTER)" Under development


Journal Side
Researcher Side
Conception
COSTER
Planning
Protocol Registration
Draft Protocol
CREST-P
Protocol Submission
TRIAGE TOOL
Review by Desk Editor
CREST-P
Peer-Review
Editorial Decision
Protocol Publication
Revision
Conduct
Reporting
Draft Manuscript
CREST
Submission
TRIAGE TOOL
Editorial Review
CREST
Peer-Review
Revision
Editorial Decision
Final Manuscript

# Implications for submitting authors

- Take advantage of our offer to review and publish protocols
- Follow best-practice standards for conduct of systematic reviews
- Think about the conduct implied by reporting standards
- For internal QC, use the same triage and peer-review tools we do
- Don’t assume that any stage of a systematic review is optional
- It’s good to be boring (results are irrelevant if methods are good)
- Find out more? **Subscribe to our newsletter**: http://bit.ly/overcite

# Thank you.

Paul Whaley | p.whaley@lancaster.ac.uk

# Image rights (in order of use)

Smokestacks / Guy Gorek / Flickr / CC BY-NC-ND

Plastic bottles / zhrefch / Flickr / public domain

Pesticide application / Oregon State University / Flickr / CC BY-SA

Food cans / King of Hearts / Wikimedia Commons

Till receipt / Till Dettmering / Wikimedia Commons

Drinks bottles / Amraepowell / Wikimedia Commons

Injected egg / Ekem / Wikimedia Commons